

Euskahaldun: euskararen aldeko martxa baten sare sozialetako islaren bilketa eta analisisia

Arkaitz Zubiaga

University of Warwick

arkaitz@zubiaga.org

Jasoa: 2015-05-28

Onartua:

Laburpena: *Gutxi dira sare sozialetan oinarrituz euskara landu duten ikerketa-lanak, eta are gutxiago Euskal Herrian ospatutako ekitaldiek sare sozialetan utzitako aztarnak aztertu dituztenak. Hutsune hori bete eta arlo honetan ikerketa sustatzeko asmoz, lan aitzindaria aurkeztea du helburu artikulu honek. Horretarako, “Euskahaldun” lemapean 2015eko Korrika martxak Twitter sare sozialean sortutako jarduna batzeko erabili dugun metodologia azaldu eta emaitza aztertzen dugu artikulu honetan. Gure analisiak erakusten duenez, zirrar handieneko momentuak Twitterren ere islatzen dira, txio kopuru handien bidez. Horrez gain, euskal komunitatean ikusgarritasuna lortu eta informazioa lau haizetara zabaltzeko ekitaldiarekin lotutako erabiltzaile kontu ofiziala eskaintzearen garrantzia erakusten dugu, eta baita kazetari eta komunikabideen parte-hartzearen beharra ere. Guztion eskura jarri ditugu Twitterrekin antzeko analisiak egiteko tresnak, antzeko ikerketa-lanak sustatzeko asmoz.*

Hitz-gakoak: sare sozialak, ekitaldiak, Twitter, jarrera, datu meatzaritza

Abstract: This work is motivated by the dearth of research that deals with social media content created from the Basque Country or written in Basque language. While social fingerprints during events have been analysed in numerous other locations and languages, this article aims to fill this gap so as to initiate a much-needed research area within the Basque scientific community. To this end, we describe the methodology we followed to collect tweets posted during Korrika, the quintessential exhibition race in support of the Basque language. We also present the results of the analysis of these tweets. Our analysis shows that the most eventful moments lead to spikes in tweeting activity, producing more tweets. Furthermore, we emphasise the importance of having an official account for the event in question, which helps improve the visibility of the event in the social network as well as the dissemination of information to

the Basque community. Along with the official account, journalists and news organisations play a crucial role in the diffusion of information. In order to encourage others to perform further research in the field, we make all the tools publicly available.

Keywords: social media, events, Twitter, behaviour, data mining

1. Sarrera

Interneten garapenarekin batera Web 2.0 fenomearen garrantzia areagotu duten tresna berriak garatu eta argitaratu dira. Izan ere, Web 2.0 delakoan erabiltzaileek, Internetetik informazioa jasotzeaz gain, modu errazean egin dezakete ekarpena eta sareko informazio-jarioa eta ezagutza aberastu. Web 2.0 tresnen artean, sare sozialak dira nabarmendu daitezkeen tresna horietako bat. Sare sozialetan erabiltzaileek parte hartu eta ekarpenak egiteaz gain, lagun-sare bat sortzeko aukera dute, eta ondorioz, aukeratutako lagun-sarearekin informazioa, argazkiak eta bideoak konpartitu, eztabaidatu eta solasteko aukera izaten dute.

Gaur egungo sare sozial nagusien artean, Twitter¹ dugu ikerketarako erabiliena. Izan ere, sare sozial honetan erabiltzaileek idazten dituzten mezuak eskuratzeko aukera eskaintzen zaigu, eta horrek aukera paregabea eskaintzen du tresna informatikoen bidez datu-bildumak batu eta ikertu ahal izateko. Beste sare sozial batzuetan ez bezala, Twitterren gehienez 140 karaktereko mezuak bidaltzera mugatuta daude erabiltzaileak. Mezu labur hauek 'txio' izena dute, eta interesgarri iruditzen zaion orok zabaldu egin ditzake; beste norbaiten txioa zabaltze horri 'bertxio' deritzo. Erabilgarritasunaren aldetik abantaila paregabeak eskaintzen ditu horrek. Besteak beste, oso erraza da edonondik sakeleko telefonoa erabiliz zure begien aurrean ikusten ari zarenaren gainean txiokatzea. Horrek, esate baterako, herri-kazetaritzaren hazkundera ahalbidetu du, edonork informa baitezake edonondik, eta ikusgarritasuna lortu Twitter bezalako sare handi batean. Era berean, ekitaldi bat ospatzen ari den bitartean, erabiltzaileek iritziak, argazkiak, bideoak eta albisteak zabal ditzakete.

Twitterrek datuok biltzeko eskaintzen dizkigun aukerei esker, interes handia piztu du komunitate zientifikoan, ikerketa egiteko informazio-iturri aberatsa baita. Ondorioz, asko dira ekitaldi entzutetsuek Twitterren utzitako islaren ikerketa egin duten lanak, New Yorkeko Occupy Wall Street [1, 2] mugimendua, Arabiar Udaberria [3] edo Munduko Futbol Txapelketa [4, 5] bezalakoak analizatuz. Baina ekitaldi horiek guztiak erdaraz eta nazioartean gertatutakoak dira, eta orain arte ez da antzeko lanik egin euskaraz eta Euskal Herrian gertatutako ekitaldiekin.

¹<http://twitter.com/>

Hortaz, lan aitzindaria aurkezten dugu artikulu honetan, euskal komunitate zientifiko arlo honetan lan gehiago burutzera eta Euskal Herriko ekitaldiak aztertzerantz animatu nahian. Hemen aurkezten den lana IkerGazte kongresuan² emandako “*Sare sozialetatik erazutako datuetan oinarritutako ikerketa*” tailerraren prestaketaren emaitza da. Twitter iturri gisa hartuta eta ikerkuntza helburu nagusi dugula, datuak biltzeko metodologia azaltzen dugu, horretarako behar diren tresnak guztion eskura jarritz. Euskararen kasua aztertzeko aukera ezinhobea eskaintzen duen ekitaldi baten adibidea aurkezten dugu artikuluan, Korrikarena³ hain zuzen ere, haren analisia egin eta emaitzak azalduz.

Metodologia hau definituz eta kodea guztion eskura jarritz, euskal komunitate zientifikoak arlo honetan ikertzen jarraitzea dugu helburu nagusi lan honekin, euskarazko eta Euskal Herriko ekitaldiek sare sozialetan sortzen duten isla aztertuz euskal komunitatearen jarrera hobeto ulertzeko.

2. Twitterren gaineko ikerketa

Twitter sare sozial nagusienetakoa bilakatu da gaur egun. Twitterrek berak ematen dituen estatistiken arabera⁴, 300 milioi erabiltzailetik gora dira hilero ekarpenen bat egingen dutenak, eta egunero 500 milioi txioren inguruan jasotzen ditu sare sozialak. Zaila da zehatz-mehatz txio hauen guztien artean euskarazkoak zenbat diren jakitea, baina Umap-ek⁵, euskarazko jarduna hein handi batean batzen duen tresnak, ematen ditu gutxi gorabeherako datu batzuk. Horien arabera⁶, 7.000 erabiltzaile baino gehiago dira nagusiki euskaraz txiokatzen aritzen direnak, eta 2015eko apirilean 226.000 txio inguru ziren euskarazkoak, egunean 7.539 batez beste. Horrenbesteko datu-jario etengabea duen informazio-iturriak ikerketarako bilduma paregabea eskaintzen du zalantzarik gabe, eta asko dira horri etekina ateratzen ari direnak.

Azken urteotan, Twitter sare sozialean hedatzen diren txioen gainean egindako ikerketa lanak nabarmen hazten ari dira. Sare soziala 2006ko martxoan⁷ abiatu bazen ere, 2010. urtean argitaratu ziren lehen ikerketa-lanak [6]. Orduz geroztik, etengabe hazten doa Twitter sare sozialean oinarritzen diren ikerketa-lanen kopurua. Izan ere, Twitter sare sozialak datuak eskuratu ahal izateko aukera paregabeak eskaintzen dizkigu ikertzaileoi, doan eta nahiko erraz lor baitaitezke datu bilduma handiak. Twitterrek hiru abantaila nagusi eskaintzen dizkigu ikerketa dugunean helburu:

- **Datu-bilduma handiak lortzeko erraztasuna:** Twitter APIa baliatuz datuak eskuratzea doakoa da, eta muga batzuk dituen arren (esate baterako, eskaera kopurua mugatua du, eta ez du datu-base osoaren %1 baino gehiago eskuratzen uzten),

²<http://www.ueu.eus/ikergazte/>

³<http://www.korrika.eus/>

⁴<https://about.twitter.com/company>

⁵<http://umap.eu/>

⁶http://umap.eu/media/pdf/umap_2015_4.pdf

⁷<https://about.twitter.com/milestones>

datu-bilduma handiak lortzeko aukera eskaintzen du, eta normalean muga horiek ez dira oztopo izaten, datu-bilduma erraldoiak nahi baldin ez badira behintzat.

- **Datuen naturaltasuna:** beste ikerketa-metodo tradizional batzuekin alderatuta, Twitterretik jasotako datuak naturalagoak dira. Metodo tradizionalan parte-hartzaileak batu ohi dira, eta ikerketaren helburua azaldu ondoren, haien erantzunak zein bestelako datuak gordetzen dira. Horrelako kasuetan gerta litekeena da parte-hartzaileek emandako erantzunak guztiz egiazkoak ez izatea, eta, ondorioz, ikerketaren emaitza neurri batean gutxienez baliogabetzea. Twitter bezalako sistemen abantaila da erabiltzaileek modu naturalean idazten dutela, nahi dutelako, eta inork behartu gabe.
- **Mota ezberdinetako datuak lor daitezke:** Askotan Twitter lagunarteko elkarrizketetarako eta “orain zer egiten ari naizen” bezalakoak zabaltzeko baino erabiltzen ez dela uste den arren, beste mota askotako informazioa topa daiteke bertan. Erabiltzaile askok eta askok inguruko ekitaldiez hitz egin ohi dute, eta batzuetan, azken orduko albisteak gertatu ahala, bertan dauden lekukoek informazio eskusiboa ere eman izan dute, tresnaren ezaugarriei esker.

Goiko abantaileri erreparatuz ikerketarako informazio-iturri perfektua dirudien arren, baditu kontuan izan beharreko hainbat desabantaila ere:

- **Datuak eskuratzen azkar ibili beharra:** Twitter APIa doakoa den arren, muga nagusienetakoa da datu historikoak eskuratzeko zailtasuna. Twitterren APIa bereziki prestatuta dago denbora errealean konpartitzen diren txioak edo azken astean bidali diren txioak eskuratzeko, eta txio zaharragoak eskuratu nahi ditugunean arazoak izan ditzakegu askotan. Txioek Twitterren datu-basean jarraitzen duten arren, ez daude modu errazean eskuragarri. Horrelakoetan, ordainpeko zerbitzuetara jotzea gomendatzen da; baina hori oztopo handia izan daiteke askotan ikertzaileontzako. Ondorioz, adi egotea eta txioak denbora errealean edo berehala, ordu gutxi batzuen buruan, eskuratzea da egokiena.
- **Hizkuntza informala:** Twitterren partekatzen den edukia aztertzean ez dute ohiko analisi linguistikorako tresnek ondo funtzionatzen. Arazo nagusia da sare sozialetan hizkuntza informala erabili ohi dela, eta, beraz, jarduna akats ortografikoez eta laburdurez josita egon ohi dela. Hori dela eta, zailagoa da edukia aztertzea, azken urteotan alor horretako ikerketa hobetzen ari den arren.
- **Gezurrak:** Albisteak jarraitzean, esate baterako, kontu handiz hartu behar izaten da Twitterren irakurritakoa, oso erraza baita albisteak asmatu eta gezurrezko istorioak sarean jartzea. Horrek ikerketan ere eragina izan dezake [7], eta, beraz, kontuz ibili behar da sare sozialetako edukiak aztertzean.

- **Datu demografikoen gabezia:** Twitterreko erabiltzaileak aztertzean, litekeena da haien datu demografikoak erabili nahi izatea batzuetan, lagina hobeto aukeratzeko edo laginaren ezaugarri demografikoak jakiteko. Zoritxarrez, baina, Twitterreko erabiltzaileek ez dute datu demografikorik argitaratzen beren profilean, eta ezinezkoa da erabiltzaile baten adina, sexua, eta abar jakitea. Kokapena jakin ahal izaten da batzuetan, baina kokapenari dagokion eremuan erabiltzaileak edozer jar dezakeenez, ez da oso fidagarria.
- **Demografikoki adierazgarri ez izatea:** Twitterren dauden erabiltzaileek ez dute gizartearen banaketa demografikoa guztiz islatzen. Sloan et al.-ek [8] erakutsi zuten, Twitterren dauden erabiltzaileak gazteak dira batez ere, eta haien profilak ez du gizartearekin guztiz bat egiten.

Desabantaila horiek guztiak kontuan izan beharrekoak dira, batez ere Twitterretik datuak biltzeko lagina ondo aukeratzuz, datuen bilketan alborapenik egon ez dadin. Horrez gain, euskararekin lan egitean aipatu beharreko beste desabantaila bat da Twitterrek eskaintzen duen hizkuntza iragazkia; izan ere, posible da txioak batzerakoan zein hizkuntzakoak jaso nahi diren zehaztea. Zoritxarrez, ordea, Twitterrek gaur egun ez du euskara antzemateko gaitasunik, eta, beraz, ezin zaio esan euskarazko txioak bakarrik jaso nahi ditugula. Oztopo horiek alde batera utzi eta ahal den neurrian saihestuta, eta abantailak erreparatuz, ordea, sare sozialen euskarazko erabileraz dagoen ikerketa areagotu egin behar dela uste dugu.

Lan honetan aurkezten den analisiari dagokionez, gainera, goiko desabantaila horiek ez dute eragin handirik, eta etorkizunean landu beharreko erronka gisa uzten ditugu. Era berean, hemen aurkezten dugun hurbilketa beste hizkuntza batzuetan eta, oro har, komunitate zientifikoan onartuta dagoena da. Artikulu honetan, Korrikaren gainean zabaldutako txioen gaineko analisia eginez, ez dugu gizartea islatzen duenik aldarrikatu nahi; beste ikuspuntu batetik, euskarazko ekitaldi batek sare sozialetan eta bereziki Twitterren sortzen duen isla aztertu nahi dugu, gaur egun sareko presentziak adierazten baitu, hein handi batean, komunitate baten eta ekitaldi zehatz baten ikusgarritasuna.

Ekitaldiak Twitter iturri gisa hartuta ikertzeari dagokionez, lan gehienak ingelesez idatzi dira, eta ingelesezko txioak aztertu dituzte [1, 2, 3, 4, 5]. Beste hizkuntza batzuetan, gaztelaniaz [9] zein portugesez [10] esate baterako, argitaratu da ikerketa-lanen bat. Euskaraz, ordea, hutsunea sumatu dugu arlo honetan, eta horixe betetzea da lan honen helburu nagusia. Bizkarguenagak [11], esate baterako, ikuspegi soziologikotik aztertu zuen gazteek sare sozialetan nola sortzen duten euskal identitatea, Facebook sare soziala aztertuz. Lan horrek, ordea, analisi kualitatiboa eskaintzen du gizarte-zientzien ikuspegi horretatik, beste ikuspegi teknikoago batetik eskain daitezkeen analisi kuantitatiboari gehiegi erreparatu gabe.

Euskaraz dugun hutsunea bete nahian, lan honek Twitter sare sozialean ekitaldi baten

gainean zabaldutako mezuak modu kuantitatiboan aztertzeke behar den metodologia definitzen du, beharrezko tresnak ikertzaileen eskura jarritz, eta 2015ean ospatutako Korrika ekitaldiaren gaineko analisisa aurkeztuz.

3. Datu bilketa

Bi erronka nagusi ditu Twitter sare sozialean zabaltzen diren mezuak jarraitzeak eta ikerketa egin ahal izateko artxibatzeak: datuak batuko dituen tresna prestatzea, eta, Twitterren jarraitu nahi den hitz eta traol sorta prestatu ondoren, datu-bilketa egitea.

3.1. Datuak biltzeko tresna

Twitterren APIak bi zerbitzu ezberdin eskaintzen ditu: (1) REST APIa, azken egunetako txioak eskuratzeko, eta (2) streaming APIa, txioak denbora errealean eskuratzeko. Kasu honetan bigarrena erabili dugu, streaming APIa alegia; izan ere, hainbat egun irauten duen ekitaldi baten jarraipena egiteko, errazagoa izaten da aukera hori erabiltzea.

Python programazio-lengoaia baliatu dugu datu-bilketarako. Lengoaia honek badu txioak batzeko pakete bat, lana asko errazten duena, *tweepy*⁸ izenekoak. Twitterretik txioak eskuratu eta gordetzeko garatutako kodea *github*-en aurki daiteke⁹. Hori erabiltzeko, eman beharreko lehen pausoa Twitterren aplikazioa sortzea da, <http://apps.twitter.com/> helbidetan. Ondoren, Twitterrek emango dizkigun lau kodeak *twitter.ini* fitxategian kopiatu behar dira. Hori egin ondoren, *tweetcollector.py* programaren bitartez txioak batzen hasi ahal izango dugu, honako komandoa erabiliz:

```
python tweetcollector.py bilketa-mota ekitaldi-izena bilaketa-katea
```

Non:

- *bilketa-mota* hauetako bat izan daitekeen: 'search-recent' (azken txioak eskuratzeko), 'search-popular' (azken asteko txio garrantzitsuenak eskuratzeko) edo 'stream' (txioak denbora errealean eskuratzeko).
- *ekitaldi-izena* eskuratu nahi den ekitaldiaren izena izango da. Nahi den izena aukera daiteke hemen, eta programak izen hori duen karpeta sortuko du *data* karpeta barruan, non txioak gordeko diren.
- *bilaketa-katea* Twitterren egin nahi dugun bilaketa da, hitz gakoak hutsunez banandurik.

⁸<http://www.tweepy.org/>

⁹<https://github.com/azubiaga/twitter-tools/tree/master/tweet-collection>

3.2. Datu-bilduma: Korrika 2015

Korrika ospatu zen 11 egunetan zehar egin genuen datu-bilketa, hau da, 2015eko martxoaren 19tik 29ra, *tweetcollector.py* programa baliatuz. Bilaketa-katea Korrikarekin lotutako hainbat traolekin osatu genuen, tartean *#korrika*, *#euskahaldun*, *#korrika2015*, *#korrika15* eta *#korrikazuzenean*. Horrenbestez, honako komandoarekin abiarazi genuen programa, eta 11 egun horietan zehar martxan eduki:

```
python tweetcollector.py stream korrika15 "#korrika"#euskahaldun"#korrika2015"...
```

Datu-bilketaren emaitza 38.276 txioko bilduma izan zen. Hurrengo atalean, txio-bilduma horren gaineko analisia aurkezten dugu. Datu-bilduma hau sarean argitaratu dugu¹⁰. Datu-bilduma honekin batera jaits daitekeen *korrika.json* fitxategia erabiltzen dugu analisirako iturri moduan, zeinak 38.276 lerro dituen, lerro bakoitzean txio bana, JSON formatuan. JSON formatuak dituen eremuen gaineko xehetasun gehiago Twitterren webgunean bertan aurki daiteke¹¹.

4. Analisia

Atal honetan analisia aurkezten dugu, bi zatitan banatuta: edukiaren analisia eta erabiltzaileen analisia. Analisisirako landu ditugun programak eskuragarri daude sarean github-en¹². Horiez gain, Barberák [12] garatutako beste programa batzuk ere erabili ditugu¹³.

4.1. Edukiaren analisia

Hasteko, edukiari erreparatuz egingo dugu analisia. Korrikak hamaika egunez jardun zuen Euskal Herrian zehar lasterka, eta aztertu nahiko genuke sare sozialetan utzitako isla nola sakabanatu den denboran zehar. Zein momentutan txiokatu zuen gehien jendeak? Horretarako, txio kopuruaren histograma aztertuko dugu, *histogram.py* programa baliatuz, eta honako komandoaren bitartez:

```
python histogram.py korrika.json korrika-histograma.dat h
```

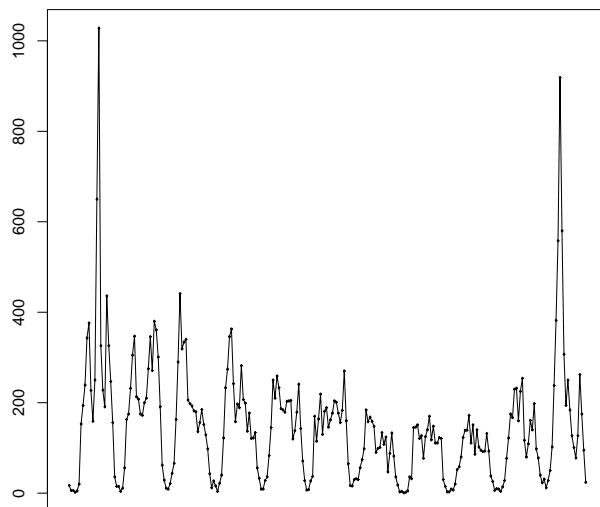
Emaitza 1. irudian ikus daiteke. Hainbat gauza erakusten ditu histograma horrek. Argi dagoena da, gauetan, Korrikak aurrera diharduen arren, txio kopuruak izugarri egiten duela behera; hori dela eta, histograman 11 egunei dagozkien igoera eta jaitsiera nabarmenak ikus daitezke. Horrez gain, lehen egunean eta azkenekoan bidali zen txio gehien; momenturik txiokatuen hasierakoa izan zen, 1000 txio baino gehiago bidali zirelarik ordubetean.

¹⁰<https://github.com/azubiaga/korrika15>

¹¹<https://dev.twitter.com/overview/api/tweets>

¹²<https://github.com/azubiaga/twitter-tools>

¹³<https://github.com/pablobarbera/pytwools>



1. irudia. Txioen histograma, orduro bidalitako txio kopurua erakusten duena.

Ondorengo egunetan, txio kopurua gutxika-gutxika jaitsi zela dirudi, Korrikaren kontaketa ez baitzen ordurako hain berria eta interesa nolabait jaitsi zelako seguruenik; eta azkenik, nabarmen egiten zuen gora berriz azkeneko egunean, Bilboko helmugara iristearekin batera.

Edukiari gertuagotik erreparatuz, interes handiena piztu zuten txioak zein izan ziren jakin nahi dugu. Horretarako, ‘bertxio’ kopurua aztertzen dugu. Zein izan ziren bertxio gehien jaso zituzten txioak? Horretarako, txio bertxiotuenen rankinga sortu dugu *top-tweets.py* programa erabiliz:

```
python top-tweets.py -v retweets -f korrika.json -k 10
```

Jarraian, 1. taulan zerrendatu ditugu Korrikan zehar bertxio gehien jaso zituzten 10 txioak. Mota ezberdinetako txioak ikus ditzakegu zerrenda horretan, baina, batez ere, ekitaldiarekin lotutako gertakizunak kontatzen dituzten txioak (adibidez, “Hasi da korrika19!”) eta txio aldarrikatzaileak (adibidez, “Korrika edo Giza katea bezalako ekimenek sortzen duten energia kolektiboa ez al dira nahikoa froga herri honen bidea zein den ohar-tzeko?”) aurki ditzakegu. Euskarazko 9 txio izateaz gain, gaztelaniazko bat ere lehen 10 hauen artean dago; izan ere, .eus Interneteko domeinuak deskontupean eskuratzeko promozio batek bertxio asko jaso zituen. Aurrez aipatu dugun bezala, momentuz, soluzio errazik gabeko erronka da Twitterren hizkuntza automatikoki antzeman ahal izatea.

Edukia beste ikuspuntu batetik begiratuta, txioak nondik bidali diren jakin nahi dugu.

Erabiltzailea (Bertxioak)	Txioa
@MeriLing1 (214)	Behin batean herri txiki handi bateko milaka pertsonak, zahar eta gaztek gau ta egun korrika egin zuten euren hizkuntzaren alde #mikroipuina
@euskaltegia (170)	Herri bat, euskaraz bizi nahian, tinko, harro. 92 urteko emakumea, Arraiotz-en. ZORAGARRIA #KorrikaZuzenean http://t.co/VsIcWoQhhM
@fm914 (149)	Ederra Indar Gorrikoek #Korrika-ri egindako agurra. Argazki andana gure guaxapean. Esker mila! http://t.co/pO3IpvLCbP
@ZuriHesian (145)	Hunkigarria da pentsatzea edozein ordutan, nonbaiten, zoro zoragarri batzuk euskararen alde #korrika dabiltzala :) http://t.co/X9OEzXoq8p
@maia_jon (132)	Korrika edo Giza katea bezalako ekimenek sortzen duten energia kolektiboa ez al dira nahikoa frogarri herri honen bidea zein den ohartzeko?
@korrika_aek (126)	Hasi da korrika19 !! Tipi tapa tipi tapa KORRIKA!!
@Hostinet (98)	Únete a la #Korrika y contrata tu #dominio .eus y #hosting web por tan solo 29,95 € con @Hostinet y @puntuus https://t.co/RTxoGkyKOn
@garesko_auzalan (96)	Korrikaren gazte kilometroa Garesen! http://t.co/HHPjaVPqJ8
@Lupilakasta (91)	Fiterotik Bilbora abia dadin Korrika!! http://t.co/w4esPCX3Xw
@EuskalakariAEK (89)	2.500 km egin ostean, lekukoa Bilbora iritsi da. Hona hemen mezua! Gora #korrika! Eskerrik asko, @katuajea! http://t.co/yB5C1FcWk4

1. taula. Txio bertxiotuenak, bertxio kopuruaren arabera ordenatuta.

Zoritxarrez, txio guztiek ez dute eskuragarri geokopakena, eta ezin da beti jakin txio bat nondik bidali den. Hori dela eta, geokopakena eskaintzen duten txioetara mugatu behar dugu ondoko analisia. Txioetatik geokopakena dutenak erauzteko, *coordinates.py* programa erabiltzen dugu behean azaldutako komandoaren bitartez, eta ondoren CartoDB¹⁴ doako zerbitzua erabiliz bistaratzen ditugu. CartoDB zerbitzuan erregistratu beharra dago, baina guztiz doakoa da, eta nahikoa da *coordinates.py* programak sortzen duen koordenatu-zerrenda igotzearekin.

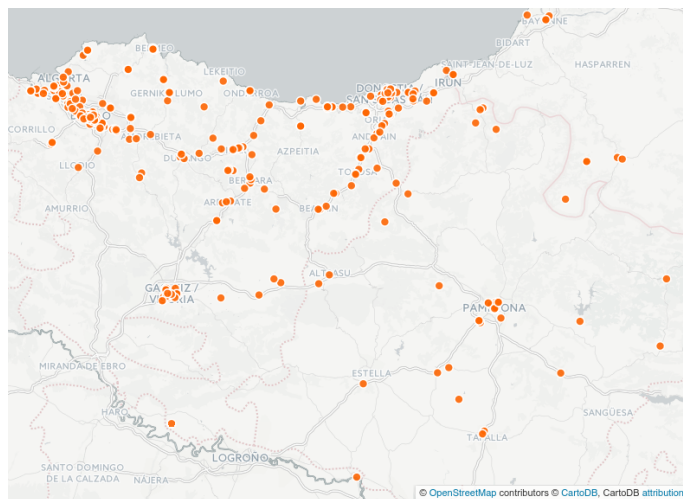
```
python coordinates.py korrika.json korrika-koordinatuak.dat
```

Emaizta bistaratuz, 2. irudiak erakusten du txio geokokatuekin sortutako mapa. Mapa horrek erakusten duenez, txioak Euskal Herri osoan zehar zabaldu ziren, lurralde guztietan. Hiriburuetan eta inguruetan txio gehiago dauden arren, Arrasate, Durango, Gernika, Tolosa, Altsasu, Lizarra, Tafalla eta beste herri askotatik ere hainbat txio bidali ziren.

4.2. Erabiltzaileen analisia

Edukia aztertu ondoren, sare sozialetan garrantzi handikoa den beste faktore bat aztertzeari ekingo diogu; erabiltzaileak, alegia. Aktiboenen zerrenda sortuz hasiko dugu erabiltzaileen analisia: nork bidali zituen txio gehien Korrikaren gainean? horretarako, *top-*

¹⁴<http://www.cartodb.com/>



2. irudia. Txio geokokatuen mapa.

tweets.py programa erabil dezakegu, komando honen bitartez:

```
python top-tweets.py -v users -f korrika.json -k 10
```

Erabiltzailea	Txio kopurua
@idorrokia	1085
@EuskalakariAEK	910
@korrika_aek	614
@HamaikaTb	444
@EAPortugalete	239
@naiz_info	233
@gaztea	227
@berria	216
@euskalirratia	208
@anabarri72	204

2. taula. Erabiltzaile aktiboenak.

Korrikan zehar txio gehien bidali zituzten 10 erabiltzaileak erakusten ditu 2. taulak. Zerrenda horretan, Korrikarekin lotutako kontuak (@EuskalakariAEK eta @korrika_aek) eta komunikabideak (@HamaikaTb, @naiz_info, @gaztea, @berria, @euskalirratia) dira nagusi, baina badaude erabiltzaile arruntak (@idorrokia, @anabarri72) eta alderdi politikoak (@EAPortugalete) kontuak ere tartean. Dena den, zerrenda horrek txio kopurua baino ez du erakusten, eta sakonago aztertu nahiko genuke nor izan diren interesgarrienak komunitatearentzat. Hori neurtzeko, bertxio kopuruari erreparatuko diogu. Zein erabiltzailek jaso zuten bertxio gehien? *top-users.py* programa erabiliz, zerrenda hau eskura

dezakegu:

```
python top-users.py korrika.json
```

Erabiltzailea	Bertxio kopurua
@EuskalakariAEK	1642
@naiz_info	918
@HamaikaTb	720
@hirual	640
@korrika_aek	568
@argia	482
@berria	385
@larbelaitz	363
@gaztea	305
@MeriLing1	243

3. taula. Erabiltzaile bertxiotuenak.

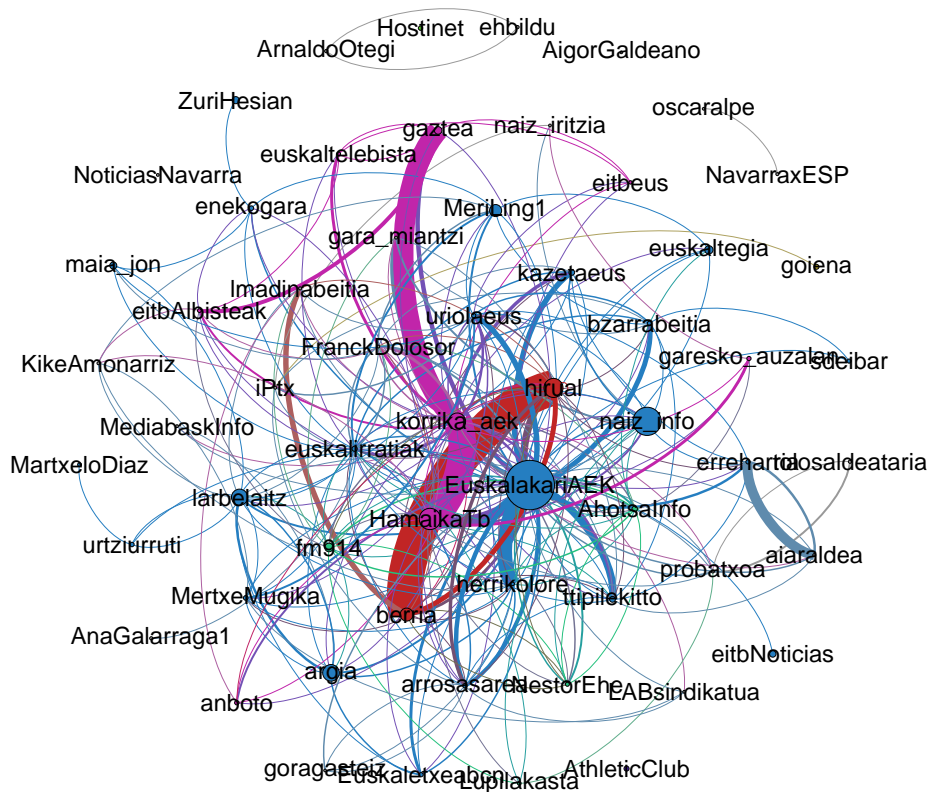
Hain zuzen, 3. taulak erakusten du bertxio gehien jaso zituzten 10 erabiltzaileen zerrenda. Lehenago azaldutako erabiltzaile aktiboekin alderatuta, badaude hor goian iraun duten batzuk, baina beste batzuk aldatu egin dira. Korrikarekin lotutako kontuak hor goian daude oraindik (@EuskalakariAEK, @korrika_aek), eta baita aurretik genituen hainbat komunikabide ere (@naiz_info, @HamaikaTb, @berria, @gaztea); baina oraingoan beste komunikabide bat ere agertu zaigu tartean (@argia), hainbeste txiokatu gabe bertxio kopuru handia lortu duena. Oraingoan, baina, erabiltzaile arruntak desagertu egin dira zerrendatik, eta hainbat kazetarik hartu dute haien lekua (@hirual, @larbelaitz, @MeriLing1). Horren arabera, badirudi kazetariak gutxiagotan txiokatu dutela, baina interes handiagoa piztuz.

Erabiltzaileen analisi honetan sakondu eta grafikoki ikusi ahal izateko, erabiltzaile nagusien sare soziala bistaratu nahi dugu. Horretarako, erabiltzaileen arteko interakzioak erauziko ditugu lehenengo. Hau da, noren txioak bertxiotu edo erantzun ditu erabiltzaile bakoitzak? Horretarako, @a erabiltzaileak @b erabiltzailearen txio bat bertxiotzen badu, interakzio moduan hartuko dugu kontuan. Era berean, @a erabiltzaileak @b-ren txio bati erantzuten badio. Korrikako txio guztietatik interakzio horien zerrenda erauzteko, *interactions.py* erabiliko dugu, honako komandoarekin:

```
python interactions.py korrika.json korrika-interakzioak.csv
```

Ondoren, interakzioen fitxategi hau bistaratzeko Gephi¹⁵ beharko dugu. Gephi software librea da eta plataforma ezberdinetarako (Windows, Linux, Mac) dago eskuragarri. Sortu

¹⁵<http://gephi.github.io/>



3. irudia. Erabiltzaile esanguratsuenekin osatutako grafoa.

berri dugun *korrika-interakzioak.csv* fitxategia Gephi-n kargatu dezakegu orain, eta sare soziala bistaratu.

Emaitza 3. irudian ikus daiteke. Erabiltzaile nabarmenenak bistaratzen dira bertan, eta beraien arteko loturaren sendotasunak interakzio kopurua adierazten du; koloreek, berriz, komunitateak. Grafo horrek @EuskalakariAEK erabiltzaile kontu ofizialaren garrantzia nabarmentzen du, Korrikaren ikusgarritasunerako eta informazioaren zabalpenerako ezinbestekoa izan baitzen. Etorkizunean ospatuko diren ekitaldientzako lezioa ere ematen digu, erabiltzaile kontu ofizial bat edukitzea zenbateraino den garrantzitsua erakusten baitu. Erabiltzaile kontu ofizialaz gain, irudiak erakusten digu inguruko erabiltzaileak gehienbat kazetariak eta komunikabideak direla. Datu horrek erakusten du ekitaldiaren ikusgarritasuna areagotzeko eta informazioa zabaltzeko lanetan zer garrantzi duten kazetariak eta komunikabideak. Horiez gain, beste erabiltzaile mota batzuen garrantzia ere ikus dezakegu, tartean futbol taldeak (@sdeibar, @AthleticClub), musikariak (@ZuriHesian) eta alderdi politikoak (@ehbildu).

5. Ondorioak

Artikulu honetan Twitter sare sozialean mundu errealeko ekitaldi batek utzitako isla ikertzeko metodologia definitu, azaldu eta praktikan jarri dugu, 2015eko Korrika aztertuz. Orain arte ez dugu antzeko analisirik ikusi euskaraz; hori dela eta, ezinbestekoa iruditzen zaigu honelako analisi bat, sare sozialen analisiak beste hizkuntza batzuetan bereganatu duen garrantzia euskarara hurbiltzeko. Analisia egiteko tresna guztiak eskuragarri daude sarean, nahi duenak ekitaldi gehiago ikertu ahal izan ditzan eta ikerketa-arlo honetan sakondu ahal izan dezan.

Korrikaren analisia bi zatitan banatu dugu: edukiaren analisia eta erabiltzaileen analisia. Biek ere aurkikuntza interesgarriak egiteko aukera eskaini digute. Edukiari dagokionez, txio kopuruak jasaten dituen gora-beherak azaldu ditugu, non une garrantzitsuenak txio kopuruan islatzen diren (kasu honetan, txio gehiago sortzen dira ekitaldiaren hasiera eta amaieran). Erabiltzaileei dagokienez, ordea, nabarmendu egin dugu erabiltzaile kontu ofiziala izateak ekitaldiaren zabalkunderako duen garrantzia. Erabiltzaile kontu ofizialaz gain, kazetariak eta komunikabideak ezinbestekoak dira ikusgarritasuna areagotu eta informazioa komunitatean zabaltzeko. Horiek guztiak kontuan izatea oso garrantzitsua da ekitaldi bat Twitter sare sozialean ikusgarri egin eta erabiltzaileen parte-hartzea sustatzeko.

Lan honen motibazio eta helburu nagusia izan da, Korrikaren kasu zehatza azalduz euskal komunitatea ulertzen laguntzeaz gain, euskararen inguruko analisi gehiago sustatzea sare sozialak ikertzeko metodologia eta tresnak partekatuz.

6. Etorkizunerako erronkak

Hemen aurkeztutako analisia hastapeneko baina ez da, eta etorkizunean antzeko analisi gehiago egin ahal izatea da orain erronka nagusia. Etorkizunean euskal ikertzaileek antzeko ikerketa-lanak gehiago lantzea eta analisi gehigarriak eginez euskarazko eta Euskal Herriko gertakizun eta ekitaldiek sare sozialetan sortzen duten islaren gainean gehiago jakitea da helburua.

Metodologia zehaztu eta tresnak liberatzeaz gain, euskarazko ekitaldien sare sozialetako isla aztertu eta ulertzeko lan asko dago egiteko oraindik ere. Alde batetik, euskararen inguruko ikerketa-lan gehiago burutzea, eta bestalde, sare sozialek dakartzaten erronkak, lan honetan laburbilduak, gainditu eta horrelako ikerketa-lanak are gehiago sendotzea. Twitterrek ez duenez ematen batutako txio guztien artean euskarazko txioak zein diren jakiteko aukera, hizkuntza-identifikazio hori egin dezakeen tresna lantzea da etorkizunerako ezinbesteko beharretako bat. Aurretik, horrelako tresna bat lantzeko asmoz, tartean Euskal Herriko Unibertsitatearekin eta Elhuyarrekin landutako TweetLID atazan [13] egin genituen lehen saiakerak eta anotatutako datu-bilduma eskaini genuen, ikertzaileek euskarazko txioen identifikazioa gehiago landu zezaten.

Bibliografia

- [1] Michael D Conover, Emilio Ferrara, Filippo Menczer eta Alessandro Flammini. 2013. «The digital evolution of occupy wall street». *PLOS ONE*, 8(5):e64679.
- [2] Benjamin Gleason. 2013. «#occupy wall street: Exploring informal learning about a social movement on twitter». *American Behavioral Scientist*, 1–17.
- [3] Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney eta Ian Pearce. 2011. «The arab spring, the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions». *International journal of communication*, 5:31.
- [4] Daniel Godfrey, Caley Johns, Carl Meyer, Shaina Race eta Carol Sadek. 2014. «A case study in text mining: Interpreting twitter data from world cup tweets». *arXiv preprint arXiv:1408.5427*.
- [5] Yang Yu eta Xiao Wang. 2015. «World cup 2014 in the twitter world: A big data analysis of sentiments in us sports fans tweets». *Computers in Human Behavior*, 48:392–400.
- [6] Haewoon Kwak, Changhyun Lee, Hosung Park eta Sue Moon. 2010. «What is twitter, a social network or a news media?» In *Proceedings of the World Wide Web Conference*, pages 591–600. ACM.

- [7] Arkaitz Zubiaga eta Heng Ji. 2014. «Tweet, but verify: Epistemic study of information verification on twitter». *Social Network Analysis and Mining*, 4(1):1–12.
- [8] Luke Sloan, Jeffrey Morgan, Pete Burnap eta Matthew Williams. 2015. «Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data». *PLoS ONE*, DOI, 10.
- [9] Eva María Ferreras Rodríguez. 2011. «Redes sociales y cambio social: el movimiento 15-m y su evolución en twitter». *Telos: Cuadernos de comunicación e innovación*, (89):61–73.
- [10] Raquel da Cunha Recuero, Gabriela Zago eta Marco Toledo Bastos. 2014. «O discurso dos #protestosbr: análise de conteúdo do twitter». *Galáxia. Revista do Programa de Pós-Graduação em Comunicação e Semiótica. ISSN 1982-2553*, 14(28).
- [11] Koldo Díaz Bizkarguenaga 2012. «Euskal identitatearen garapena online sare sozialen bidez: gazteen praktika sozialak facebooken». *Gogoa*, 11(1).
- [12] Pablo Barberá. 2015. «Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data». *Political Analysis*, 23(1):76–91.
- [13] Arkaitz Zubiaga, Iñaki San Vicente, Pablo Gamallo, José Ramon Pichel, Iñaki Alegria, Nora Aranberri, Aitzol Ezeiza eta Víctor Fresno. 2014. «Overview of tweetlid: Tweet language identification at SEPLN 2014». *TweetLID @ SEPLN*.